Glycoengineering in CHO cells: Advances in systems biology

Tejwani, Vijay; Andersen, Mikael Rørdam; Nam, Jong Hyun; Sharfstein, Susan T

Published in:
Biotechnology Journal

Link to article, DOI:
10.1002/biot.201700234

Publication date:
2018

Document Version
Peer reviewed version

Review

Glycoengineering in CHO cells: Advances in systems biology[†]

Vijay Tejwani[1], Mikael R. Andersen[2], Jong Hyun Nam[3], and Susan T. Sharfstein[1]

[1] Colleges of Nanoscale Science and Engineering, SUNY Polytechnic Institute, Albany, NY USA

[2] Department of Biotechnology and Biomedicine, Technical University of Denmark, Lyngby, Denmark

[3] Export-Import Bank of Korea, Seoul, South Korea

Correspondence: Professor Susan Sharfstein, Colleges of Nanoscale Science and Engineering, SUNY Polytechnic Institute, 257 Fuller Road, Albany, NY 12203 USA

E-mail: ssharfstein@sunypoly.edu

Keywords: cellular engineering, Chinese hamster ovary cells, glycosaminoglycans, mathematical modeling, protein glycosylation

[†]This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: [10.1002/biot.201700234].

Received: August 17, 2017 / Revised: December 28, 2017 / Accepted: December 29, 2017

Abstract

For several decades, glycoprotein biologics have been successfully produced from Chinese hamster ovary (CHO) cells. The therapeutic efficacy and potency of glycoprotein biologics are often dictated by their post translational modifications, particularly glycosylation, which unlike protein synthesis, is a non-templated process. Consequently, both native and recombinant glycoprotein production generate heterogeneous mixtures containing variable amounts of different glycoforms. Stability, potency, plasma half-life, and immunogenicity of the glycoprotein biologic are directly influenced by the glycoforms. Recently, CHO cells have also been explored for production of therapeutic glycosaminoglycans (e.g. heparin), which presents similar challenges as producing glycoproteins biologics. Approaches to controlling heterogeneity in CHO cells and directing the biosynthetic process toward desired glycoforms are not well understood. A systems biology approach combining different technologies is needed for complete understanding of the molecular processes accounting for this variability and to open up new venues in cell line development. In this review, we describe several advances in genetic manipulation, modeling, and glycan and glycoprotein analysis that together will provide new strategies for glycoengineering of CHO cells with desired or enhanced glycosylation capabilities.

Abbreviations: 3OST1, Heparan sulfate 3-O-sulfotransferase 1; **ADCC,** Antibody-dependent cell-mediated cytotoxicity; **AGL,** α-Glucosidase; **alpha-Gal,** Galactose-α1,3-galactose; **CDC,** Complement-dependent cytotoxicity; **CMP,** Cytidine monophosphate; **ER α-Man,** Endoplasmic reticulum mannosidase; **Fuc,** Fucose; **FucT,** Fucosyltransferase; **GAG,** Glycosaminoglycan; **Gal,** Galactose; **GalNAc,** N-acetylgalactosamine; **GalNAz,** N-azido-galactosamine; **GalT,** Galactosyltransferase; **GDP,** Guanosine diphosphate; **GlcA,** Glucuronic acid; **GlcNAc,** N-acetylglucosamine; **GnT,** *N*-acetylglucosaminyltransferases; **hCG,** Human chorionic gonadotropin hormone; **IdoA,** Iduronic acid; **IFN-γ,** Interferon gamma; **LacNAc,** N-acetyllactosamine; **mAB,** Monoclonal antibody; **Man,** Mannose; **Man-II,** Golgi-mannosidase II; **NDST2,** N-deacetylase/N-sulfotransferase 2; **Neu5Ac,** N-acetylneuraminic acid; **Neu5Gc,** N-glycolylneuraminic acid; **NeuX,** Sialidase X; **NSD,** Nucleotide sugar donor; **OST,** Oligosaccharyltransferase; **PFR,** Plug flow reactor; **P-P-Dol,** Dolichol pyrophosphate; **rhEPO,** Recombinant human erythropoietin; **rHuAChE,** Recombinant human acetylcholinesterase; **SA,** Sialic acid; **SAS,** Sialic acid synthetase; **SAT,** Sialic acid transporter; **t-PA,** Tissue plasminogen activator; **UDP,** Uridine diphosphate; **α-2,3-SiaT,** α-2,3-sialyltransferase; **α-2,6-SiaT,** α-2,6-sialyltransferase; **α-Man,** α-Mannosidase

1. Introduction

Chinese hamster ovary (CHO) cells are the most widely used host for manufacturing complex biopharmaceuticals due to their ability to replicate folding and post-translational modifications, including glycosylation patterns, found in human proteins [1, 2]. CHO cells also grow robustly in suspension in serum-free media and have a long history of regulatory approval and an established track record for producing safe, efficacious products [1, 3, 4]. In 1987, the US Food and Drug Administration (FDA) approved the first CHO-derived recombinant protein, tissue plasminogen activator (t-PA) for use as a therapeutic [5]. Since this approval, CHO cells have been the clearly preferred choice for commercial production of glycoprotein biologics, ranging from antibodies to hormones to cytokines [6]. In 2016, 5 of the top 10 biopharmaceuticals were produced in CHO cells [7] while 10 out of the 15 biopharmaceuticals approved by the FDA in 2016 were produced in CHO cells, including 3 biosimilars [8].

Correct glycan structures are crucial for potency and control of pharmacokinetic and pharmacodynamic properties of glycoprotein biologics and therapeutic carbohydrates [2, 9]. Protein glycosylation and synthesis of the glycosaminoglycan (GAG) portion of proteoglycans are non-templated, and thus, significant heterogeneity can arise from organism to organism, cell type to cell type, and even between different culture conditions. While CHO cells produce the most "human-like" glycans of all rodent cell lines, immunogenic epitopes such as galactose-α1,3-galactose (alpha-Gal) or N-glycolylneuraminic acid (Neu5Gc) can occur as terminal units on glycans [10],[11]. Glycoprotein biologics produced from CHO cells generally contain low amounts of Neu5Gc and may or may not contain the alpha-Gal epitope. Thus, recombinant glycoproteins are considered to be safe for use in humans as these immunogenic epitopes are generally present in a very low amounts, insufficient to elicit an immune response, but since CHO cells contain enzymes to produce these epitopes in the glycans, monitoring biologics for their presence is important [12],[13]. Additionally, CHO cells lack the enzymes, α-2,6-sialyltransferase (α-2,6-SiaT) and β-1,4-*N*-acetylglucosaminyltransferase (GnT-III), which are responsible for generation of terminally linked α-2,6-sialic acids and bisecting N-acetylglucosamine (GlcNAc) addition, respectively, in human glycan structures [14-16]. α-2,6-linked sialic acids and bisecting GlcNAc can be important for potency of glycoprotein biologics [17-20]. Several studies have demonstrated that recombinant proteins with desired glycosylation profiles have higher stability, potency, half-life in blood circulation, and reduced immunogenicity in comparison to their wild type counterparts [21-26].

From over three decades of research, mostly involving labor-intensive and time-consuming empirical processes, the volumetric productivities of recombinant protein from CHO cells have increased from 0.05 g/L to >10 g/L [27, 28]. Despite this progress, there is a huge gap in understanding the molecular basis of protein and carbohydrate production in CHO cells. The recent publication of genome sequences for CHO-K1 [29], Chinese hamster [30], and six additional CHO cell lines [31] provides a starting framework for deeper understanding of key cellular processes like transgene expression, metabolism and protein secretion, which drive recombinant protein expression in CHO cells. Indeed, availability of the CHO genome sequence has accelerated research on different "omic" fronts including genomics, transcriptomics, glycomics, proteomics, and metabolomics [28, 32-35]. The next steps in this area are to combine the enormous volume of molecular process data generated from different omics technologies using mathematical/computational models to provide holistic and system-level understanding of cellular physiology in CHO cells [35]. Complete understanding of such cellular processes in CHO cells is vital for metabolic bottleneck identification and for rational development of next-generation CHO cells with novel features in a time-saving-manner. Different omics technologies advancing CHO cell biotechnology are described in recent reviews [34-36]. In this review, we describe

different strategies and advances in glycoengineering of recombinant proteins and production of GAGs in CHO cells. We first describe the process of glycosylation followed by different strategies for controlling or enhancing glycosylation in recombinant proteins produced in CHO cells. We then review different mathematical models whose predictive functions have been of great value in cellular engineering of CHO cells rationally to achieve desired glycosylation profiles in glycoprotein biologics. Finally, we discuss analytical advances and examples of how system-wide analysis of glycoprofiles provides insight into biological processes in CHO cells and guidance for novel strategies to improve glycan profiles.

2. Protein Glycosylation

Carbohydrate modification falls into three general categories. N-linked glycans are attached to the amide nitrogen in the asparagine side chain, O-linked glycans to the oxygen in the side chains of serine, threonine, hydroxylysine or hydroxyproline, and glycosylphosphatidyl inositol (GPI) derivatization occurs on the carboxy-terminal carboxyl group. In addition, there are some unusual forms of glycosylation recently discovered. A glycoprotein may have one or more glycosylation sites. Each putative glycosylation site may be always occupied, variably occupied (referred to as macroheterogeneity) or never occupied. In addition, the glycan attached to each site may vary from molecule to molecule (referred to as microheterogeneity), leading to a highly variable population of glycoproteins from a single polypeptide backbone.

2.1 N-linked glycosylation

In the case of N-linked glycans, the consensus sequence Asn-X-Ser/Thr where X is any amino acid except proline is necessary, but not sufficient, for glycosylation. Between 10% and 30% of potential glycosylation sites are not occupied [37]. N-linked glycans can be grouped into three main categories, all of which contain a common core structure ($Man_3GlcNAc_2$); complex, which contain no mannose (Man) residues other than in the core, high mannose, and hybrid, in which some of the branches are of the complex type and some of the high mannose type (Figure 1A). In each case, the GlcNAc is linked in a β-linkage to the amide nitrogen of the asparagine side chain. In the complex structures, each branch frequently terminates in a sialic acid although sulfated lactosamine is also seen, particularly in glycoprotein hormones. Site analysis has shown that the distribution of different classes of N-linked structures is frequently specific for each site on a protein. For example, in the rat brain protein Thy1, site 23 has only high mannose structures; site 74 has only complex and hybrid structures, and site 98 contains all three types [38, 39].

Proteins targeted for secretion or membrane insertion are translocated to the endoplasmic reticulum by means of a 15-30 amino acid signal sequence comprised of primarily hydrophobic amino acids during translation. The newly synthesized protein enters the lumen of the endoplasmic reticulum where it is subject to co-translational folding and N-linked glycosylation (Figure 2). N-linked glycosylation occurs by the transfer of a lipid-linked oligosaccharide species ($Glc_3Man_9GlcNAc_2$-PP-dolichol) to an asparagine residue on the newly synthesized polypeptide chain. The oligosaccharide is trimmed by two glycosidases and one or more mannosidases in the endoplasmic reticulum and the protein is then translocated to the Golgi apparatus where it undergoes further oligosaccharide chains modification and O-glycosylation, followed by packing and processing for secretion.

2.2 O-linked glycosylation

The most common form of O-linked glycosylation is the mucin type, which contains an α-glycosidic linkage between N-acetylgalactosamine (GalNAc) (and sometimes GlcNAc) and the hydroxyl group of serine or threonine (Figure 1B). There is no clear consensus sequence for O-linked glycosylation; however, the serine or threonine acceptor amino acid appears to be preferentially surrounded by

serine, threonine, proline, alanine, and glycine residues [40]. O-linked glycosylation of secreted proteins occurs in the Golgi apparatus and hence, is limited to residues present on the protein surface after the protein is completely folded in the ER. Currently, there are eight different core structures known to occur beyond the Ser/Thr-GlcNAc linkage (Figure 1B). In CHO cells, core 1 glycans dominate the O-linked glycan structures [41]. There are a limited number of monosaccharides present in mucin-type glycans, namely GalNAc, galactose (Gal), GlcNAc, fucose (Fuc), and sialic acids. Sulfation is also a frequent modification. In general, glucose (Glc) and Man are not found in these structures. Biosynthesis occurs in a stepwise fashion with individual nucleotide-activated monosaccharides added by site-specific glycosyltransferases. While the O-linked mucin structures are much more compact and generally smaller than the N-linked structures, they are also much more diverse. For example, human lung mucin contains more than 100 different glycans [40, 42]. A fairly common structure, particularly for cell-surface glycoproteins, is the presence of many O-linked glycan in close proximity. These structures appear to serve both as passive and active barriers preventing proteolytic and infectious attacks.

2.3 Proteoglycans

Proteoglycans consist of a glycosaminoglycan attached to a core protein, through a β-linkage between xylose and either serine or threonine on the core protein. GAGs are long (up to ~1 MDa) linear chains of repeating disaccharide units, consisting of either GlcNAc or GalNAc alternating with glucuronic acid (GlcA) and/or iduronic acid (IdoA) or Gal. Glycosaminoglycans fall into four categories, the heparin/heparan sulfate GAGs, which consist of repeating units of GlcA or IdoA and GlcNAc , chondroitin sulfate/dermatan sulfate GAGs, which consist of repeating units of GlcA or IdoA and GalNAc, keratin sulfate GAGs, which consists of repeating units of GlcNAc and Gal, and hyaluronic acid GAGs, which consist of repeating units of GlcA and GlcNAc. Keratin sulfate can be N-linked to asparagine via a β-linkage to GlcNAc or O-linked to Ser/Thr through an α-linkage to GalNAc [43]. Hyaluronic acid is unique in that it is not attached to a core protein and is not sulfated. In contrast, all the other GAGs may contain sulfate modifications on more than 50% of the disaccharides in the GAG chain.

3. Manipulating glycosylation and glycosaminoglycan production by genetic approaches

The cellular machinery of CHO cells synthesizes complex-type N-glycans as heterogeneous mixtures of bi-, tri- and tetra-antennary structures containing varying amount of Fuc, GlcNAc, Gal, Neu5Gc and Neu5Ac on the $Man_3GlcNAc_2$ core [44] as shown in Figure 2. N-linked glycosylation strongly affects therapeutic activity and efficacy of glycoprotein biologics. For example, monoclonal antibodies (mABs) devoid of core Fuc at Asn^{297} in the Fc region showed up to 50-fold increase in antibody-dependent cell-mediated cytotoxicity (ADCC) [45, 46]. In another study, high mannose content in the Fc region of a mAB significantly reduced its *in vivo* half-life [47]. Thus, manipulating N-glycosylation to increase desired glycans in glycoprotein biologics is of high commercial interest**.** Genetic approaches for enhancing N-glycosylation are the most common and employ gene editing or transient expression techniques to change the activity of glycosyltransferases and increase or decrease the precursors involved in the N-glycosylation process. Genetic approaches altering heterogeneity, sialylation, fucosylation and branching in N-glycan structures are described below. Further, engineering approaches to increase putative N-glycosylation sites on the protein of interest, examples of biopharmaceuticals with O-linked glycans, and strategies to improve productivity and potency of GAGs are also discussed in this section. A summary of genetic approaches for manipulating glycosylation in CHO cells is illustrated in Figure 3.

3.1 Manipulating heterogeneity

Although glycoprotein biologics produced by CHO cells containing heterogeneous mixtures of N-glycans are considered safe as human therapeutics, excess heterogeneity can be an issue. This excess

heterogeneity arises due to the variability of N-glycan processing and can compromise the safety and activity of such glycotherapeutics. Having homogenous glycoforms allows comparative studies of their biological effects, which can be advantageous in development of therapeutic candidates [9]. Yang et al. showed that CHO cells can be genetically engineered to produce glycoproteins in a nearly homogenous form without any deleterious effect on their growth or other compensatory changes [44]. To achieve this goal, the *in vivo* function of each of the nineteen glycosyltransferases potentially participating in N-glycan formation and processing was determined in CHO cells by individual and/or multiple glycosyltransferase gene knockouts. The effects of knocking out each of the 19 genes involved in N-glycan branching (*mgat1/2/3/4A/4B/4C/5/5B*), galactosylation (*B4galt1/2/3/4*), N-acetyllactosamine (LacNAc) elongation (*B3gnt1/2/8*), terminal capping by sialylation (*st3gal3/4/6*) and core α-6-fucosylation (*fut8*) were determined using a stably expressed model protein, a recombinant human erythropoietin (rhEPO) containing in its structure three N-glycans with heterogeneous tetra-antennary structures, low poly-LacNAc and terminal α-2,3-linked sialic acid. From the 19 glyco-genes, gene knockouts in CHO cells identified to be influential in decisive steps of N-glycosylation process are described below.

The knockout screens showed that the model glycoprotein, rhEPO devoid of β4-branched tetra-antennary structures, can be produced in CHO cells by combined knockout of *N*-acetylglucosaminyltransferases (GnT) genes, *mgat4A* and *mgat4B*. Similarly, rhEPO devoid of β6-branched tetra-antennary structures can be produced by knockout of *mgat5* gene. Combined knockout of *mgat4A/4B/5* genes enables the CHO cells to synthesize rhEPO containing mostly homogenous bi-antennary *N*-glycans (with a small amount of poly-LacNAc). The galactosylation in the model glycoprotein was reduced by over 90% by combined knockout of galactosyltransferase (GalT) genes, *B4galt1/3,* with *B4galt1* knockout being the major contributor for the observed decrease in GalT activity in CHO cells. Knockout of *B4galt1* in CHO cells containing no *mgat4A/4B/5* genes eliminated galactosylation in the homogenous bi-antennary *N*-glycans. Poly-LacNAc addition, a poorly understood process, could be eliminated from rhEPO by *B3gnt2* gene knockout. Similarly, sialylation was removed from rhEPO by stacked knockout of sialyltransferases genes, *st3gal4/6*. Further, the combination of *st3gal4/6* and *mgat4A/4B/5* knockouts in CHO cells produced bi-antennary N-glycans lacking sialylation with increased poly-LacNAc. Knockout of fucosylation gene, *fut8,* encoding α-1,6-fucosyltransferase (FucT) resulted in rhEPO lacking core Fuc in the N-glycan structure. In the same study, combined knockout of *fut8* and *B4galt1* genes in CHO cells expressing a recombinant IgG1 resulted in homogeneous bi-antennary *N*-glycans lacking Fuc, containing minor amounts of Gal. Other individual gene knockouts (*mgat3, mgat4C, mgat5B, B3gnt1, B3gnt8 and st3gal3*) or combination of genes (*B4galt1/2/4*) showed no effect on the respective steps in the N-glycosylation process in CHO cells [44].

3.2 Manipulating sialylation

Sialylation refers to the glycosidic addition of a negatively charged monosaccharide, a sialic acid, by sialyltransferases, generally to terminal Gal or GalNAc though occasionally to GlcNAc or sialic acid itself in complex N-glycans [48]. CHO cells contain α-2,3-sialyltransferases (α-2,3-SiaT), whereas humans cells have α-2,6-SiaT in addition to α-2,3-SiaT [14, 15]. Due to this difference, N-glycans produced by CHO cells only contain sialic acid residues linked by α-2,3-glycosidic linkages, whereas human glycans contain both α-2,3- and α-2,6-linked sialic acid residues [49, 50]. Sialic acid at the termini in complex N-glycans masks terminal Gal from recognition by hepatocyte asialoglycoprotein receptors that leads to rapid clearance from the circulation [21]. Insufficient or lack of sialylation in glycoprotein biologics can lead to inconsistency in the pharmacodynamics and cause challenges in formulating reproducible dosages. Thus, correct and generally, maximal sialylation is necessary to ensure longer plasma half-lives and maximum *in vivo* activity and therapeutic efficacy [51].

Of the numerous (>50) sialic acids occurring in nature, Neu5Ac is the most abundant [48]. Its derivative, Neu5Gc, is also a major sialic acid found on mammalian cell surfaces, formed by Neu5Ac hydroxylation by cytidine monophosphate (CMP)-Neu5Ac hydroxylase. Humans lack CMP-Neu5Ac hydroxylase and are unable to synthesize Neu5Gc. Glycoproteins synthesized by CHO cells occasionally contain N-glycans capped with Neu5Gc. Neu5Gc in glycans can be a cause for concern as Neu5Gc-capped glycans act as "xenoautoantigens" in humans and are cause of "xenosialitis", an inflammatory process initiated by binding of naturally occurring antibodies against Neu5Gc in the human body [52-54]. The Neu5Gc-dependent antigenicity of glycoproteins obtained from CHO cells depends on the amount and locations of Neu5Gc in the glycan structure. For example, mABs and rhEPO produced by CHO cells containing 1-2% Neu5Gc did not elicit an immune response, whereas in the same study, feutin with high levels of Neu5Gc (7% of total sialic acid residues) elicited an immune response in chickens [12]. A recently study by Yu et al. on different clinical mAbs containing Neu5Gc residues concluded that mABs containing a single Neu5Gc residue do not bind to anti-Neu5Gc antibodies, while only a minor fraction of mABs containing two or more Neu5Gc showed binding to anti-Neu5Gc antibodies [55]. Further, this study suggested that binding of anti-Neu5Gc antibodies to mAbs containing multiple Neu5Gc depends on Neu5Gc location in the mAb structure.

Glycoprotein biologics that contain α-2,6-linked sialic acid residues are suggested to be more "human like" and perform better *in vivo* [17, 56]. In two recent studies, chemo-enzymatic modification (*in vitro* glycosylation) of two IgGs (anti-Her2 antibody and Rituximab) creating homogenous glycans containing α-2,6-sialic acid residues enhanced ADCC due to stronger interaction of α-2,6-sialylated glycans with FcγRIIIa receptors on natural killer cells [17, 18]. Similarly, in two previous studies involving *in vitro* glycosylation, α-2,6-sialylated versions of IgGs produced superior anti-inflammatory responses compared to unsialylated or α-2,3-sialylated versions [56, 57]. Genetic strategies to introduce α-2,6-sialylation and/or increase terminal sialic acid residues are discussed below.

3.2.1 Increasing α-2, 6 sialylation

CHO cells were successfully engineered to produce rhEPO containing almost exclusively α-2,6-sialylation by knockout of *st3gal4/6* genes (encoding α-2,3-SiaTs) and knock-in of *st6gal-I* gene (encoding an α-2,6-SiaT) [44]. Furthermore, in the same study, homogeneous bi-antennary N-glycans capped by α-2,6-NeuA were produced by additional knockout of *mgat4A/4B/5* genes. In a similar study, Chung et al. achieved successful α-2,6-sialylated IgG by amino acids substitution in the IgG structure to allow access by the sialyltransferase enzyme. When expressed in CHO cells containing *st3gal* gene knockout and knock-in of *st6gal-1* gene, IgG with four amino acid substitutions contained nearly exclusively α-2,6-sialylation with a 14-fold increase in sialic acid content compared to wild type IgG obtained from unmodified CHO cells [58]. While combining gene editing techniques with protein engineering for optimal α-2,6-sialylation may be necessary, effects of mutations in a mAb should be carefully considered as amino acid substitutions can elicit an anti-inflammatory response as well as decreasing the antibody affinity towards the Fcγ receptor, leading to a decrease in ADCC [59, 60].

3.2.2 Increasing the sialic acid content

Besides knockout and knock-in of sialyltransferase genes, increases in sialic acid content can be obtained by overexpressing genes participating in steps prior to sialylation or inhibiting genes encoding for sialidases that remove the sialic acid after the sialylation step.

3.2.2.1 Overexpression of genes prior to the sialylation step.

3.2.2.1.1 Overexpression of galactosyltransferase genes

Addition of sialic acid residues to a growing N-glycan chain can be limited by the absence of Gal, which acts as an acceptor substrate for sialyltransferases. Increasing the Gal content in N-glycans by

overexpressing GalT produces a corresponding increase in the sialic acid content in CHO cells overexpressing sialyltransferase enzymes. Raymond et al. transiently co-expressed genes encoding for GalT, α-2,6-SiaT, and an IgG1 antibody (trastuzumab/Herceptin®) and produced efficient α-2,6-sialylation in trastuzumab's Fc region. The glycans under investigation were monosialylated, a physiologically relevant form found in circulating human IgGs. Over 85% of sialic acids on trastuzumab showed α-2,6-sialylation due to overexpression of human GalT, which inserts Gal residues in the glycan that are preferentially used by α-2,6-SiaT rather than α-2,3-SiaT. Overexpression of GalT alone increased the Gal content of the Fc glycans, but had no effect on increasing the sialylation, whereas overexpression of α-2,6-SiaT only increased the sialylation moderately [61]. In a similar study, co-expression of GalT was beneficial in CHO cells overexpressing α-2,3-SiaT to increase the sialylation. Tri-sialylated glycans on rhEPO increased from 17.3% to 35.5% when expressed in CHO EC1 cells co-overexpressing both human α-2,3-SiaT and GalT [62].

3.2.2.1.2 Overexpression of anti-apoptotic genes

If during a bioprocess, cell death and consequent cell lysis is minimized, concentrations of sialidases in the extracellular medium will be reduced during prolonged cell-culture experiments. Overexpression of anti-apoptotic proteins like Bcl-X_L and 30Kc19 increases culture longevity by suppressing apoptosis, leading to increased sialylation due to reduced extracellular sialidase activity [63, 64]. CHO cells overexpressing *30Kc19* gene not only increased rhEPO sialylation by 87.1 % but also increased its yield by 102.6 % [64].

3.2.2.1.2 Overexpressing genes involved in sialic acid biosynthesis and transport

Sialic acid content can be enhanced by increasing the concentration and availability of the donor sugar nucleotide CMP-sialic acid (CMP-SA) present in the Golgi apparatus. Inside the nucleus, CMP-SA is generated from sialic acid by CMP sialic acid synthetase (CMP-SAS) and later transported to the Golgi by CMP-sialic acid transporter (CMP-SAT). In eukaryotes, sialic acid is synthesized in the cytoplasm by three enzymes in a four-step process. The two first steps are catalyzed by a bifunctional enzyme, GNE (UDP-N-acetylglucosamine-2-epimerase/N-acetylmannosaminekinase), having kinase and epimerase activity. The epimerase activity of GNE converts UDP-GlcNAc to N-acetylmanosamine (ManNAc), which is then converted to ManNAc-6-phosphate by the kinase activity of GNE. The last two steps are catalyzed by two enzymes, Neu5Ac-9-phosphate synthase and Neu5Ac-9-phosphate phosphatase, which produce Neu5Ac from ManNAc-6-phosphate by condensation and dephosphorylation reactions respectively. GNE is a rate-limiting enzyme for synthesis of sialic acid in the cytoplasm [15, 65].

Increasing sialic acid content in the cytoplasm or CMP-SA in the nucleus by different strategies increases sialylation by varying amounts. Overexpression of CMP-SAT in CHO cells expressing recombinant human interferon gamma (IFN-γ) increased IFN-γ sialylation by 4–16 % [66]. Similarly, a modest increase in sialylation (10-20%) was observed upon supplementation with ManNAc, which led to a 12-fold increase in the intracellular pool of CMP-sialic acid [67]. In another study, supplementation with 1,3,4-O-Bu_3ManNAc, a chemical analog of the sialic acid precursor ManNAc increased final sialic acid content of rhEPO >40% in CHO cells at a 100-fold lower concentration than natural ManNAc [68].

Co-overexpressing several genes together enhanced sialylation more significantly than overexpressing single genes individually. CHO cells expressing CMP-SAS in combination with CMP-SAT and human α-2,3-SiaT exhibited greater rhEPO sialylation compared with CHO cells overexpressing α-2,3-SiaT or CMP-SAS individually [69]. GNE, catalyzing the first two steps in the synthesis of sialic acid, is a rate-limiting enzyme, and its activity is regulated by feedback inhibition from free cytoplasmic CMP-Neu5Ac [15, 65]. Co-expressing a mutant version of GNE lacking feedback regulation with CMP-SAS enhanced the sialylation modestly, but when the mutant version of GNE was co-expressed with CMP-SAT and human α-2,3-SiaT, CHO cells produced rhEPO having a 43% increase in sialylation. Co-expression experiments involving CMP-SAT clearly indicated that endogenous CMP-SAT is insufficient, and its overexpression was essential for increasing the sialylation [70].

3.2.2.2 Inhibition of genes after the sialylation step

Sialidases are enzymes that catalyze the removal of sialic acid residues from glycoproteins and glycolipids. CHO cells contain four different sialidases (Neu1-4) distributed in the lysosome (Neu1 and Neu4), cytosol (Neu2), and plasma membrane (Neu3) [71]. During glycoprotein biologic manufacture, lowering the activity of these enzymes is desirable, but not completely, due to their important roles in crucial biological functions [72-75]. The cytosolic sialidiase, Neu2 is released into the supernatant during cells lysis and preferentially removes α-2,3-linked sialic acids from the glycoprotein products [76]. RNA-mediated suppression of Neu2 [77] and plasma membrane-bound Neu3 [51] activities by 40% and 98%, respectively, resulted in increased sialic acid content (up to 33% in model proteins) but the effect of Neu2 suppression was observed in death phase only.

3.3 Manipulating fucosylation

Core Fuc residues on IgG antibodies have a negative effect on their effector function [45]. Effector function is essential in IgGs designed for use in tumor therapy; after binding to antigens on cancer cells, the IgG Fc region binds strongly to FcγRIIIa receptors present on natural killer cells, causes cancer cell death by lysis via the ADCC mechanism [45, 78]. Fucose-containing N-glycans present at Asn^{297} in the IgG adversely affect the Fc-FcγRIII interaction [79]. Multiple studies have successfully demonstrated that removal of the core Fuc residue from the N-glycan in human IgG_1 increases the binding affinity of Fc towards FcγRIII, which, in turn, enhances the *in vivo* ADCC significantly [45, 78]. Such fucose-free antibodies can be beneficial to patients as their higher potency enables administration of lower dosages [80, 81].

In mammals, *fut8* is the only gene encoding for a fucosyltranferase capable of adding Fuc to N-glycans [82]. Inhibition or knockout of *fut8* and interference with transport and synthesis of donor substrate GDP-Fuc are two approaches for reducing or inhibiting Fuc addition to N-glycans in CHO cells [83]. Knockout of *fut8* in CHO cells using zinc finger nucleases and homologous recombination produced fully afucosylated antibodies, showing enhanced ADCC [84, 85]. Further, Chan et al. showed that inactivating *Slc35c1*, encoding for GDP-fucose transporter, also generated CHO cells producing fucose-free glycans [86]. *Slc35c1* was inactivated separately by ZFNs, TALENs, and the CRISPR-Cas9 techniques. Mutant CHO cells produced by these three different knockout techniques produced a model glycoprotein, EPO-Fc fusion protein and an IgG_1 (Herceptin) completely devoid of core Fuc residues on their N-glycans. These mutations showed no negative effects on cell growth, viable cell density, or antibody productivity [86].

Alternatively, antibodies devoid of core Fuc can also be produced by CHO cells overexpressing the GnT-III enzyme, which catalyzes bisecting GlcNAc addition onto the common core structure ($Man_3GlcNAc_2$) in N-glycans [16]. Upon addition of bisecting GlcNAc, the oligosaccharide cannot act as a suitable substrate for subsequent glycosylation enzymes, especially Golgi-mannosidase II (Man-II), GalT, and FucT [87], leading to production of fucose-free mAbs showing increased ADCC, but decreased complement-dependent cytotoxicity (CDC). The decreased CDC is due to hybrid N-Linked oligosaccharide structures, resulting from incomplete mannose cleavage [88]. In this study and others, CDC could be increased to normal or even higher levels by co-expression and defined spatial localization of Man-II in combination with overexpression of GnT-III [16, 88]. Modulating the expression of glycosyltransferase enzymes and engineering their localization in the Golgi thus serves as a powerful way for production of tailored glycoengineered therapeutic antibodies with enhanced effector and complement function.

3.4 Manipulating branching

In mammals, branching in N-glycans occurs in the medial Golgi and is carried out by GnT-I, -II, -IV and –V enzymes, which control GlcNAc addition at the branch point in a stepwise manner [87]. Branching produces bi-, tri-, and tetra-antennary structures, which can be extended by enzymes in the trans-Golgi capable of adding Gal, Fuc, and sialic acids (Figure 2) [89]. GnT-I and GnT-II control the formation of bi-antennary structures; tri-antennary structure formation is controlled by GnT-IV or GnT-V, and the tetra-antennary structure formation is controlled by the combined action of GnT-IV and GnT-V (Figure 2). Branched structures in N-glycans are involved in physiologically important events such as cellular proliferation and signaling as well as pathological conditions including tumor progression and metastasis [90-93]. Higher branching (tri and tetra) provides additional sites for attachment of sialic acid residues, which enhance biological activity and circulatory lifetime. Increased branching of therapeutic glycoproteins is thus of significant clinical as well as commercial interest.

Multiple studies have shown that branching can be increased by overexpression of GnT-IV and/or GnT-V enzymes. Tri-antennary structures were significantly increased (over 50%) by overexpression of GnT-IV or GnT-V individually in CHO cells producing IFN-γ with predominantly bi-antennary sugar chains [94]. In the same study, tetra-antennary structures were increased up to 56% of the total sugar chains when GnT-IV and GnT-V enzymes were co-expressed. However, compared to observed increases in tri- and tetra-antennary structures, the corresponding increase in sialylation was insignificant. Inadequate sialylation was attributed to insufficient intracellular sialyltransferase or CMP-SAS activity. The reduced sialic acid capping on tri- and tetra-antennary structures in rhEPO was solved by Yin and coworkers by additionally expressing human α-2,6-SiaT [95]. They showed that co-expression of α-2,6-SiaT with GnT-IV and GnT-V produced rhEPO containing approximately 92% tri- and tetra-antennary N-glycans with a 45 % increase in the sialic acid content compared with rhEPO obtained from wild-type CHO-K1 cells.

3.5 Engineering the protein of interest

An alternative route to optimizing glycan structures on recombinant glycoproteins is engineering or modifying the protein itself for attachment of desired or additional N-glycans. As described in Section 2.1, in the N-glycosylation process, the N-glycan is attached to the asparagine residue in the consensus tripeptide Asn-X-Ser/Thr present in the protein structure, where X can be any amino acid except proline [96-98]. Putative N-glycosylation sites can be introduced in a protein by engineering in the Asn-X-Ser/Thr tripeptide. Glycoprotein biologics with desired N-glycan structures or in a hyperglycosylated form capped with sialic acid residues produced using such protein modification/engineering techniques often have a longer half-life and enhanced biological activity.

When incorporating additional N-glycan sites, the bulkiness of the negatively charged glycan can interfere with the binding property of the glycoprotein [99]. The N-glycan addition site should be chosen away from the glycoprotein binding site and located in a protein structure accessible to the the oligosaccharyltransferase enzyme complex involved in the transfer of N-glycans from the lipid-linked complex to the asparagine in the nascent polypeptide [100]. Glycosylation efficiency not only depends on the accessibility of the N-glycan site but also on the type of amino acid occupying the middle position in the Asn-X-Ser/Thr tripeptide [101-104], the presence of serine or threonine [103], and the type of amino acids present near the tripeptide [105-107]. Proline is completely disfavored at X while leucine, negatively charged residues like glutamate and aspartate, or bulky side chains like tryptophan also show poor glycosylation efficiency [97, 100, 102, 108]. For example, threonine in the Asn-X-Ser/Thr sequence in rabies virus glycoprotein enhanced its N-linked glycosylation compared with serine in the same position [103].

Recombinant proteins like EPO and follicle-stimulating hormone engineered with additional N-linked glycosylation sites showed at least threefold enhancement in serum half-life and increased *in vivo* activity compared to wild-type versions [109, 110]. Darbepoetin alfa, a recombinant version of EPO engineered to contain two additional N-glycans besides the three naturally occurring glycosylation sites, increased patient convenience and compliance as longer serum half-life and higher potency correspond to reduced injection frequency and dosage [111]. However, additional N-glycan sites do not necessarily result in improved biological properties as vacant sialic acid attachment sites, lower oligomerization state, and exposed Gal residues in the extra N-glycans can have a negative effect on half-life [112-115]. For example, recombinant human acetylcholinesterase (rHuAChE) produced by HEK293 cells in a triglycosylated form as a mixture of dimers, trimers, and tetramers showed a serum half-life of 80 min. Hyperglycosylated forms of rHuAChE engineered to contain one or two additional N-glycan sites showed faster clearance from the blood stream due to a number of unoccupied sialic acid attachment sites in the additional two N-glycans. The half-life of undersialylated rHuAChE was increased to 19 hours from 80 min by co-administration of saturating amounts of asialofetuin, a compound that saturates the hepatic asialoglycoprotein receptors, delaying the clearance of hyposialylated glycoproteins [115]., Chitlaru et al. extended the previous finding and proved that the N-glycans in rHuAChE are not efficiently sialylated due to inefficiency of sialyltransferase enzymes in HEK293 cells. In this case, undersialylation exposed Gal residues in the N-glycan structure, marking the rHuAChE for rapid plasma clearance. Chitlaru et al. also showed that co-expression of α-2,6-SiaT in the HEK293 cell line expressing rHuAChE in a hyperglycosylated form led to almost fully sialylated glycoforms, significantly increasing the plasma half-life. The half-life of efficiently sialylated rHuAChE glycoforms was further extended by *in vitro* tetramerization of lower oligomeric forms [112, 113].

3.6 O-linked glycoengineering

To date reports of O-linked glycoengineering for glycoprotein biologics are limited, although the O-glycoproteome has been characterized for CHO-K1 cells as described in Section 5. However, a number of glycoprotein biologics contain O-linked glycans either naturally occurring or as a result of molecule design. CHO-derived biopharmaceuticals containing O-linked glycans include rhEPO [95, 116], human chorionic gonadotropin [117], human Factor VIII [118] and Factor IX [119]. Examples of Fc-fusion-protein biopharmaceuticals engineered to contain O-glycan structures include Etanercept (Enbrel®) [120], Abatacept (Orencia®) [121], Corifollitropin alfa (Elonva®) [122] and ACP-501 (an Fc-fusion with recombinant human lecithin-cholesterol acyltransferase) [123]. These biopharmaceuticals contain one to several O-glycan structures, mainly for protection from proteolytic degradation.

3.7 Glycosaminoglycan engineering

Sulfated glycosaminoglycans coat the surfaces of all living animal cells and are integral components of the extracellular matrix [124]. In addition to their biological functions, purified GAGs have a variety of medical applications including surgical aids (hyaluronic acid and chondroitin sulfate), tissue engineering (hyaluronic acid, chondroitin sulfate, and heparin/heparan sulfate) and anticoagulant activity (dermatan sulfate and heparin) [125]. By far the most important GAG (and one of the most important drugs in medicine today) is anticoagulant heparin. Heparin is the most widely used anticoagulant drug in modern medicine; approximately 300,000 doses/day are used in the U.S., and greater than 100 tons of heparin are used annually, with a market value of ~$7 billion [126, 127]. A health crisis in 2008, involving the adulteration of heparin produced from hogs in China, led to the death of ~100 Americans and resulted in a demand for heparin from non-animal sources [128]. In addition to a demand for a safer source of heparin, produced under controlled, good manufacturing process conditions, recent studies suggest that heparin may have significant antineoplastic activity, separate and distinct from its anticoagulant activity [129-133] while other studies indicate a role for heparin in treating inflammation, infertility, and infectious disease [134-138]. While heparin is produced only in mast cells, the related GAG, heparan sulfate, is expressed in all mammalian cells.

Heparan sulfate differs from heparin in that the overall sulfation level is lower, the GAG chain length is longer and heparan sulfate lacks anticoagulant activity due to the absence of an antithrombin-binding pentasaccharide sequence in the GAG chain.

In an effort to produce a bioengineered heparin, Baik and coworkers engineered CHO cells to overexpress two critical enzymes in the heparin biosynthetic pathway, N-deacetylase/N-sulfotransferase 2 (NDST2) and heparan sulfate 3-O-sulfotransferase 1 (3OST1) [139]. Through this overexpression, they were able to increase the amount of GAGs secreted into the culture medium by ~10-fold and the anticoagulant activity of the GAGs secreted into the medium by nearly 100-fold. However, the activity was still ~40-fold lower than pharmacological heparin. Of equal concern was that the structure of the engineered GAG differed substantially from pharmacological heparin. Pharmacological heparin exhibits a significant fraction of di- and tri-sulfated disaccharides, whereas in the engineered "heparin", the primary disaccharide structures contained a single sulfate modification on the amino-group of GlcNAc with very few di- or tri-sulfated structures seen. In addition, confocal microscopy indicated that the 3OST1 was not localized to the Golgi (as expected), but rather was seen distributed throughout the cells. In a subsequent study where a Golgi-targeted version of heparan sulfate 3OST1 was overexpressed without NDST2 addition, a 5-fold increase in both di- and tri-sulfated GAGs was observed compared with wild-type cells although the total amount of GAGs produced was not altered [140]. More significantly, tetrasaccharide analysis by LC/MS [141] revealed the presence of one of the five 3-O-sulfated tetrasaccharides observed in bovine lung heparin in the engineered cells, but not in the wild-type CHO cells. Despite the challenges faced in developing bioengineered GAGs, particularly heparin, this series of papers was the first to demonstrate metabolic engineering for production of a non-protein product in CHO cells.

4. Predictive N-linked glycoengineering using mathematical models for a systems biology approach.

Microheterogeneity during N-glycan processing is attributed to the complex network of enzymes participating in the glycosylation reactions and the mixing characteristics in the Golgi [142]. If the Golgi behaves like a plug-flow reactor (PFR), proteins will tend to have homogeneous glycans, as each protein will have the same residence time. In contrast, if the Golgi behaves like a stirred tank with significant mixing, residence times will vary, leading to increased heterogeneity. This phenomenon cannot be investigated by experiments alone due to the complexity of the glycosylation reactions, including enzyme localization in different compartments, an enormous network of glycosylation enzymes, substrate preferences among the enzymes, and redundancy among enzymatic activities [143]. Mathematical models provide an alternative to experimental investigations in addressing the effects of cellular or process changes introduced to increase recombinant protein productivity or maximize desired glycoform fractions within the glycoform population. As mathematical models have advanced, their predictive powers have increased in correctly guiding cellular engineering efforts, reducing the time required to develop a cell line producing desired glycosylation patterns [142-150]. The major mathematical models developed to address the N-glycosylation process and aid the cellular engineering efforts are described below in chronological order according to the year of development.

4.1 Glycosylation models (named by author initials and year)

4.1.1 UB1997 model: The first comprehensive model of N-glycosylation addressing glycoform heterogeneity was proposed by Umana and Bailey in 1997 [144]. This model is based on the vesicular transport hypothesis (awarded the Nobel Prize in 2013) and considers Golgi compartments as static containers wherein the transport of proteins in the network is approximated by four continuous well-mixed reactors in series. This model incorporated kinetic values for parameters (up to the first

galactosylation step) available or estimated from the CHO cell literature available at that time. The parameters incorporated in this model include concentrations, kinetic constants, and distribution of N-glycosylation enzymes in the different Golgi compartments. Additionally, the volume of the Golgi compartments, the specific productivity of the glycoprotein, and half-lives of proteins in the Golgi were also incorporated in the model. The UB1997 model contains 33 reactions and is capable of generating 33 different oligosaccharide structures from 8 enzymes localized in the 4 different Golgi compartments. The glycan distribution predicted by the UB1997 model for three proteins, t-PA, rhEPO and β-interferon, containing mainly complex oligosaccharides, was similar to the experimental distribution observed when these proteins were produced in CHO cells. Further, the usefulness of this model for CHO glycoengineering applications was shown by its ability to identify the parameters requiring manipulation during the design of metabolic engineering experiments aimed at optimizing a particular glycoform distribution. The UB1997 model suggested that formation of a desired glycoform fraction (complex-bisected oligosaccharides) could be significantly increased by a carefully balanced overexpression of three enzymes (GnT-III, GnT-II, and Man-II) rather than overexpression of GnT-III alone, which increased the proportion of undesirable (serum half-life lowering) bisected hybrid oligosaccharides. Bisected complex oligosaccharides can enhance antibody biological activity (e.g., Campath-1H) compared to non-bisected glycans. Predicting the need to overexpress several enzymes instead of one is difficult without the aid of such a model [144].

4.1.2 KB2005 model: Krambeck and Betenbaugh extended the UB1997 model by incorporating three additional enzymes and similarly using literature information related to CHO cells to estimate normal ranges for the additional model parameters [150]. The additional enzymes permitted the KB2005 model to analyze additional steps in N-glycan processing including fucosylation, extension of antennae by LacNAc repeats, and sialylation. This network of 11 enzymes increased the network of reactions to 22,871 and model capability to generate 7565 glycan structures. In contrast to the UB1997 model, this model contains algorithms for adjusting model parameters such as enzymatic activities to match glycoform distribution found in experimentally produced glycoproteins from CHO cells. Additionally, an algorithm for optimizing model parameters to match a desired glycoform is included in the KB2005 model. This study suggested that such model enhancements can provide a much faster way to produce a product with a desired glycoform distribution pattern. A modified version of the KB2005 model containing no GnT-III enzyme was recently shown by McDonald et al. to identify novel control points in the glycosylation process that can be modified to increase the proportion of high antennary structures in complex N-glycans [151]. The modelling experiments identified the activity of a GalT, occurring downstream of the glycan branching points, as the major determinant that governs the proportion of tri- and tetra-antennary glycans on a nascent protein; simulations also suggested that overexpression of GnT-IV and/or -V would only increase the percentage of N-glycans with tri- or tetra-antennary structures by a modest amount. The modelling results from the KB2005 model suggested that the relative proportion of tetra-antennary structures can be increased by 10-14 fold (and tri-antennary structures by 2-3 fold) by translational suppression (over 90%) of GalT4 enzyme alone. These model findings were validated experimentally in engineered CHO cells (with GnT-IV and -V overexpression and GalT4 knockdown) expressing a human chorionic gonadotropin (hCG) hormone, which is heavily glycosylated, mainly containing bi-antennary structures. The engineered cell line produced recombinant hCG with a 10-14-fold increase in tetra-antennary structures and a 2-3-fold increase in tri-antennary structures, which is in line with the 10-fold increase observed in the simulations. In contrast, cell lines overexpressing GnT-IV and -V enzymes individually or dually without the knockdown did not demonstrate a marked increase in tri- and tetra-antennary structures.

4.1.3 HH2007 model: Instead of modelling the Golgi as four continuous, well-mixed reactors in series, based on the vesicular transport model as done by Umana and Krembeck, Hossler et al. developed a mathematical model based on Golgi maturation theory [142]. This model considers Golgi

compartments as dynamic and approximates the Golgi as four PFRs. The authors argued that glycosylation reactions simulated using models resembling PFRs are more likely to be accurate compared to simulations performed using models based on vesicular transport. The applicability of this model to aid glycoengineering tasks was shown by its suggestion that all terminally processed N-glycans can be produced in a homogenous form by spatial localization of enzymes to specific compartments along with sufficient holding time. The usefulness of enzyme localization for obtaining homogenous glycans has been verified experimentally in yeast and mammalian cells [152].

4.1.4 KB2009 model: The algorithm used in the development of the KB2005 model is capable of predicting the majority of N-glycans occurring in CHO cells. The algorithm was expanded by Krambeck et al. in 2009 to analyze N-glycosylation in human cells. [149] The expanded model, KB2009, included 8 additional glycosylation enzymes for generating additional structures found in human N-glycans, has the ability to generate a theoretical mass spectrum from the corresponding glycan structure, and is able to analyze experimental MALDI TOF mass spectra of human N-glycans. These features power the KB2009 model to predict changes in activity and levels of 19 glycosylation enzymes from mass spectrometry data. Additionally, the model can predict the glycan structure along with the proportions of isomers within each peak from measured glycan mass spectra. Further, when supplied with mass spectrometry data obtained from analysis of glycans found to be different between normal and diseased cells, the model predicted the expected shift in levels of certain glycosyltransferases known to occur differentially between healthy and diseased cells. The predictive and analytic power of the KB2009 model can be useful for glycoengineering applications in CHO cells as impacts of enzyme knock-in/knockout on the glycan profile can be readily inferred. These applications were further explored in an updated 2017 version of the model [143]. In this work, the mathematical framework was updated and structurally extended to include larger network of glycan structures (50,605 structures based on a maximum mass cutoff, but can be increased as needed). The updated model (KB2017) was applied to analyze glycomic profiles of Lec and LEC CHO mutants [153]. KB2017 model was able to predict complex features of the glycosylation mutants as well as regulatory mechanisms, showing a great potential for predicting effects of glycoengineering.

4.1.5 JK2011 model: The models developed previously by Krambeck and Hossler did not account for differences in the kinetics of enzymes participating in the N-glycosylation process. Also, they assumed that the concentrations of nucleotide sugar donors (NSDs) in the Golgi are constant, which is not the case, as the concentrations of almost all NSDs inside the Golgi depend on the availability of the respective NSDs in the cytosol. The latter assumption of constant concentration isolates the Golgi from cellular metabolism, delinking the glycosylation process and cellular metabolism [148]. In reality, NSDs (UDP-GlcNAc, GDP-Fuc, and UDP-Gal) are synthesized in the cytosol and transported into the Golgi by transport proteins. Inside the Golgi, the NSDs participate in the N-linked glycosylation process by acting as building blocks for sugar moieties addition [154-156]. Jimenez del Val et al. advanced the previous models developed on Golgi maturation theory to overcome the two limitations described above. The model constructed by Jimenez del Val incorporated parameter values for 8 enzymes involved in the Golgi-linked N-glycosylation process and is capable of generating 77 structures and 95 reactions. The JK2011 model is fed with literature-derived kinetic information for individual enzymatic reactions occurring in the Golgi apparatus combined with kinetics of transport proteins moving NSDs into the Golgi. Additionally, enzymes and transport proteins with unknown concentrations were included by estimating their concentrations using optimization-based methodologies. The JK2011 model derived from these refinements was able to accurately predict the glycosylation profiles found in some commercial mABs. The JK2011 model was also able to confirm the experimental finding that NSD concentrations vary throughout the course of cell culture and that oligosaccharide distribution is sensitive to that variation [155, 157, 158]. Additionally, the JK2011 model outperformed previous models in terms of accuracy and ability to successfully replicate the experimental findings of *fut8* gene-silencing effects and effects of cytosolic NSD depletion on glycosylation patterns in mABs [148]. The

latter finding, when coupled with NSD metabolism, can directly link observed glycosylation patterns in mABs to extracellular metabolites like glucose, whose concentration can be easily quantified.

4.1.6 JK2014 model: The JK2011 model was advanced further to estimate and relate intracellular concentrations of NSDs to extracellular nutrients like glucose and glutamine by incorporation of kinetic parameters for two additional processes, cell growth and synthesis of NSDs [159]. These extensions were included to provide a link between feeding strategies and product glycoform distributions. Incorporation of NSD concentrations estimated by this model into the JK2011 model provided a glycan distribution profile in the Fc region of a mAB closely matching experimental data from a murine hybridoma cell line. Additionally, this model was able to link consumption of glucose and glutamine with cell culture dynamics (cell growth and cell death), providing for the first time, an *in-silico* platform that could be developed further for bioprocess optimization with respect to product quality.

4.1.7 SL2016 model: A common disadvantage of the computational models described above is that these models require specification of many kinetic parameters, some of which are not readily available or fully detailed, e.g., enzyme *in vivo* binding affinities. To overcome this disadvantage, Spahn et al. [147] developed a novel model based on Markov chain theory and built on a probabilistic framework, creating a low-parameter model, requiring no input of estimated (or any) kinetic parameters. Instead of creating the reaction network using kinetic information, this model requires as an input, a fit with experimentally derived, wild-type glycoform profiles from a producer cell line. The algorithm adopts enzyme reaction rules reported in literature and used in the KB2009 model. The model utility in predictive glycoengineering applications in CHO cells was shown by its ability to accurately predict the effects of gene knockouts for GnT-IV and core FucT on rhEPO and IgG, respectively. In another study, Spahn et al. [146] demonstrated that the algorithm can successfully predict the experimental perturbations required to replicate the glycosylation pattern in biosimilar versions of two glycoprotein therapeutics.

These models are summarized in Table 1.

5. Advances in analytical techniques and omics technologies

The very first genomic [29] and transcriptomic [160] studies of a CHO cell line addressed the issue of glycosylation. Xu et al. found that out of 300 human genes associated with glycan synthesis and degradation, only three genes (*ALG13*, *CHST7* and *CHST13*) lack homologs in the CHO-K1 genome. However, they observed that approximately half of the genes were not expressed, including many of the sulfotransferases, fucosyltransferases and GalNAc transferases [29]. Working independently, without access to the genome sequence, Becker et al. observed transcripts for all the relevant reactions necessary for producing complex N-glycans [160]. The only glycosylation-related functions that were not detected within the transcriptome data encode the GnT-IVa, which adds GlcNAc to Man molecules in the glycans, creating tri- and tetra-antennary glycans, and the UDP-Gal transporter (UGT) that is responsible for the transport of cytoplasm-derived UDP-Gal to the Golgi lumen. In the first major proteomic study after the publication of the CHO genome, Baycin-Hizal and coworkers analyzed lysates and secreted proteins from CHO-K1 cells [161]. To identify glycoproteins after tryptic digest, samples were oxidized with sodium periodate and mixed with hydrazide resins to bind the glycopeptides to the resin. PNGaseF treatment was used to release the peptides from the resin-bound N-linked glycans, and the peptides were analyzed by mass spectrometry. They were able to identify enriched functional groups in the glycoproteome including cellular components, binding, and catalytic activity.

A number of recent advances in glycan and glycopeptide analysis, including improved separation methods for removing glycoproteins from a total cell lysate or supernatant, advances in labeling, and improvements in liquid chromatography and mass spectrometry have enhanced our ability to analyze

both the glycans and glycopeptides for N-linked and O-linked oligosaccharides as well as glycosaminoglycans [124, 162-168]. In particular, sample enrichment is critical, as glycoproteins occur at relatively low frequency. Enrichment techniques include lectin enrichment, hydrophilic interaction chromatography, boronic chemistry, hydrazide chemistry, reductive amination chemistry, oxime click chemistry, and non-reductive amination chemistry [164]. In addition, detailed structural information, such as the stereochemistry of glycosidic linkages and branching patterns, can now be obtained from complex mixtures of glycoconjugates. As a result of these advances, we are now entering the glycomic or glycoproteomic era.

To date, a limited number of studies have examined the glycome or glycoproteome of CHO cells. Liu et al. examined the effects of O-linked glycoengineering on CHO-K1 cells [169]. CHO-K1 cells, in general, are not able to elongate or branch core 1 O-glycans (Figure 1B) because they lack the necessary glycosyltransferase activity, In addition, core 3 O-glycans are also absent in CHO-K1 cells. Glycoproteins derived from CHO-K1 cell lines are devoid of more complex terminal carbohydrate determinants such as blood group ABO, Lewis and sulfated determinants. Liu et al. transiently transfected the mucin-type fusion protein PSGL-1 (fused with mouse IgG2b) with β-1,3-N-acetylglucosaminyltransferase 3, core 2 β-1,6-N-acetylglucosaminyltransferase I, or core 3 β-1,3-N-acetylglucosaminyltransferase 6 into CHO-K1 cells. By combining western blotting with LC/MS, they were able to map the O-glycome and demonstrate extended core 1 and core 3 glycans, and increased expression of core 2 O-glycans. Yang et al. applied their SimpleCell strategy to CHO-GS cells by knocking out the Cosmc gene with a Zinc-finger nuclease [170]. Cosmc is a private ER chaperone (a chaperone dedicated to the folding or assembly of a single protein or family) required for the O-glycan core 1 synthase that catalyzes the second step in O-glycan elongation, adding β-3-Gal to the initial GalNAc residues attached to the protein backbone. The resulting O-glycans have only the initial GalNAc residue. Using these cells, they characterized the O-glycoproteome in total lysates as well as the secretome and identified a total of 738 O-glycoproteins and 1548 O-glyco-sites. Using a lectin capture strategy, they analyzed wild-type CHO cells and identified a combined total of 824 O-glycoproteins and 1727 glyco-sites between the wild-type and mutant. In a creative application of Click-chemistry (a reaction between an azide and an alkyne yielding a covalent product), Slade and coworkers characterized the secretome of O-linked glycans in CHO cells by feeding the cells N-azido-galactosamine (GalNAz) [171]. Secreted proteins labeled with GalNAz were recovered from the spent medium using an alkyne-modified agarose resin. They envision that this technique will allow for identification and characterization of host cell proteins and provide strategies for improved protein purification.

6. Conclusions and perspectives

As CHO-produced therapeutics continue to expand, combined with the development of biosimilars, interest in manipulating glycan distributions will increase apace. While genomic, transcriptomic, and proteomic studies have identified the enzymes and transporters that control glycosylation, the delicate interplay between the enzymes, substrates and the protein of interest remains to be elucidated. Together, advances in omics technologies, cell editing approaches, glycan analysis and glycosylation modeling will hopefully provide guidance for glycoengineering to achieve exquisite control of glycan structures and provide improved therapeutics with better efficacy, stability, pharmacokinetics and pharmacodynamics.

Acknowledgements

This work was supported in part by a grant from the National Science Foundation, IIP-1622959.

Conflict of Interest

The authors declare no financial or commercial conflict of interest.

Table 1: Comparison of number of enzymes, reactions and glycan structures present in different computational models [142-144, 147-150, 159]

Model (reference)	Enzymes	Enzymatic reactions	Glycan structures	Framework Theory
UB1997 (144)	8	33	33	Vesicular transport
KB2005 (150)	11	22,871	7565	Vesicular transport
HH2007 (142)	10	Not provided	329	Golgi maturation
KB2009 (149)	19	~ 40,000	10,000 – 20,000[a]	Vesicular transport
JK2011 (148)	8	95	77	Golgi maturation
JK2014 (159)	30	60	34	Golgi maturation
SL2016 (147)	10	Not provided	Not provided	Markov chain & Flux-balance
KB2017 (143)	13	19,413[a] /34,872[a]	50,605[a] /1,00,464[a]	Vesicular transport

[a] Actual number depends on the molecular weight cutoff implemented in the model.

Figure Legends

Figure 1: A. Overview of the three main types of N-glycans. B. Seven core structures of mucin-like O-glycosylation found in humans. Core 8 consists of Gal and GalNAc in an α1-3 linkage. From information found in [46].

Figure 2: A. Schematic of the modification of N-glycans in the endoplasmic reticulum and the Golgi complex. B. Sugar and enzyme abbreviations

Figure 3: Summary of genetic approaches for manipulating glycosylation in CHO cells described in this review. Gene knockout are represented by yellow segments and gene knock-in, over expression and alteration are represented by green segments. For convenience, tetra-antennary N-glycans and genes participating in N-glycosylation are also shown.

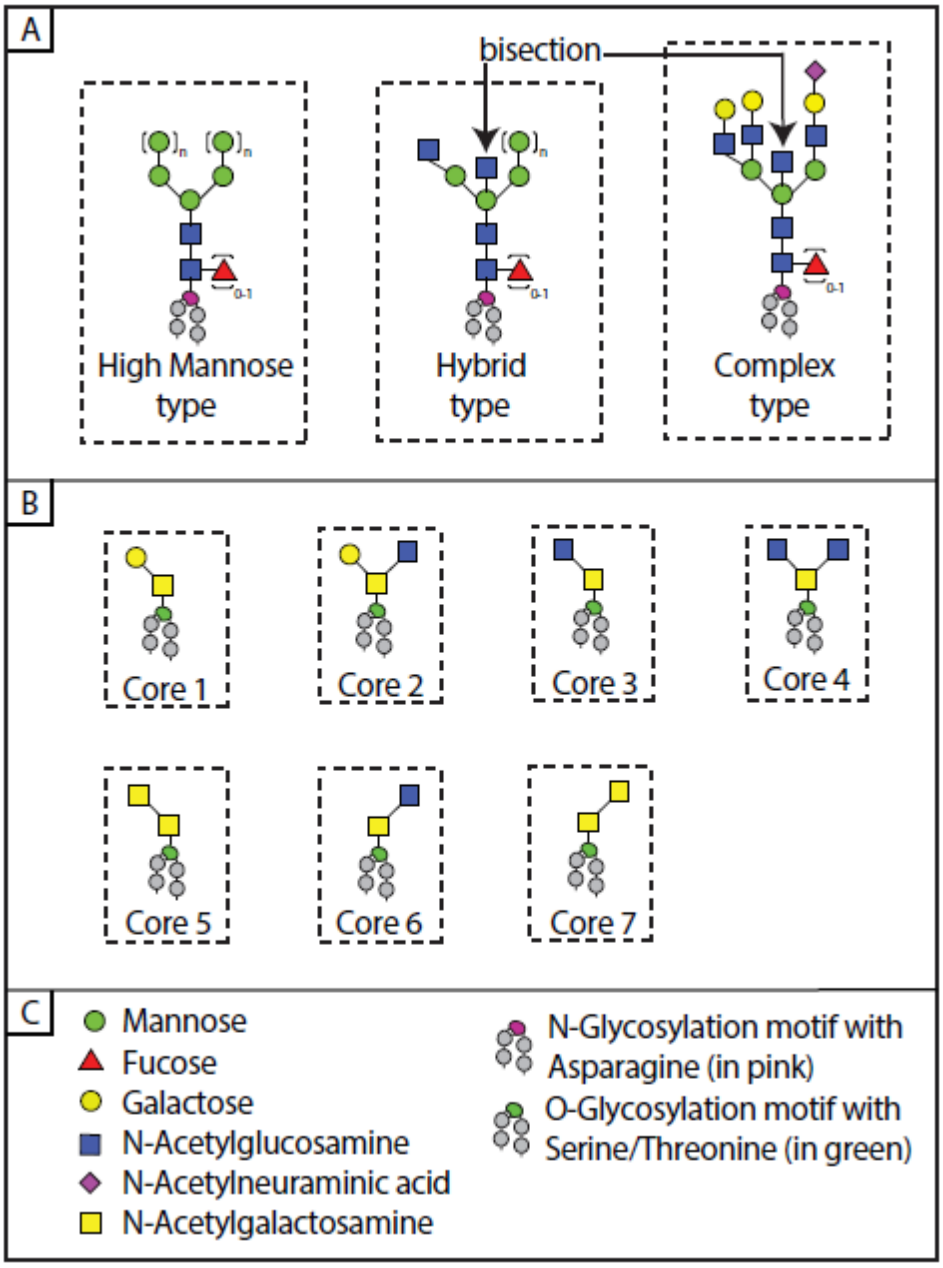

Figure 1

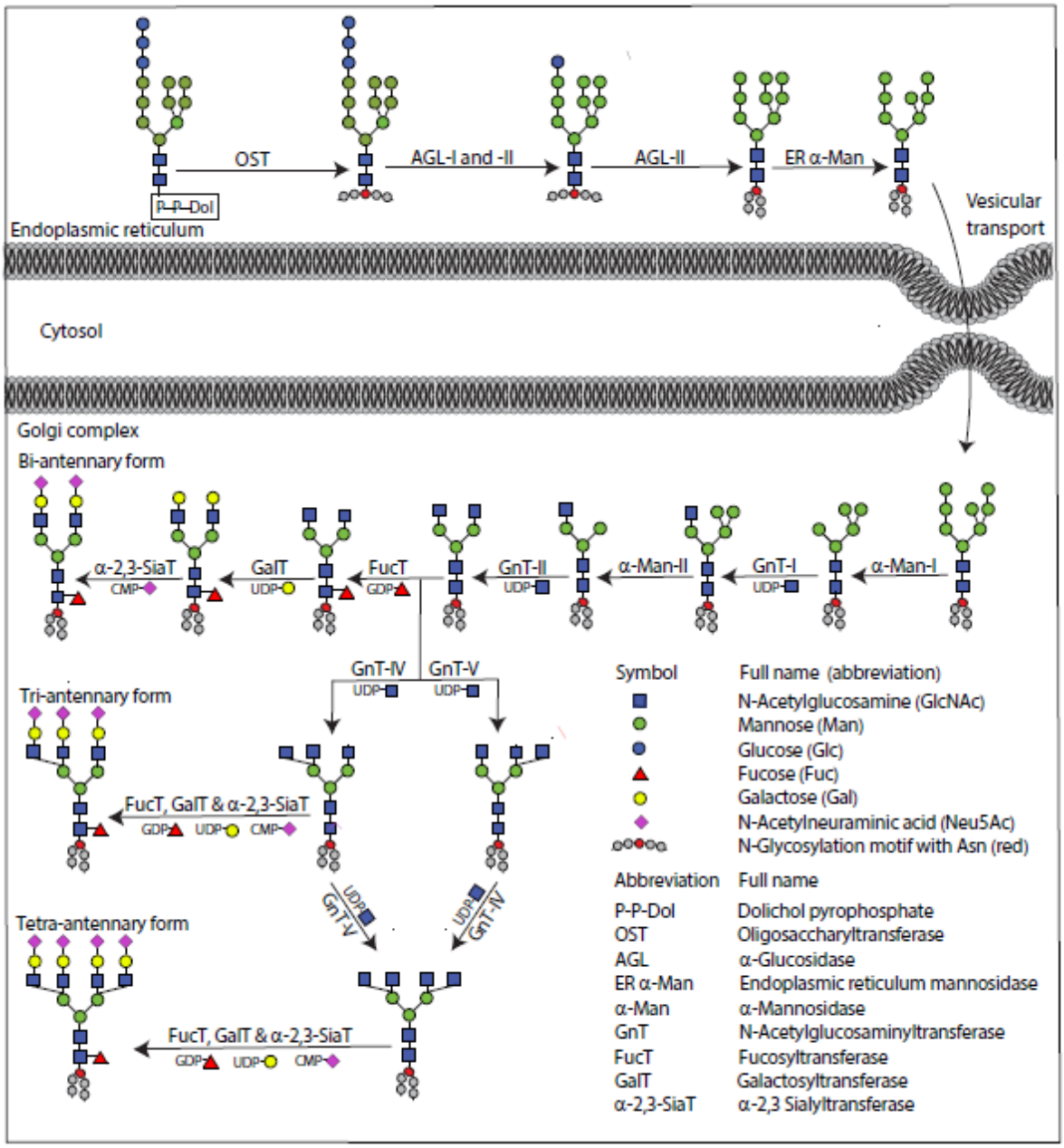

Figure 2

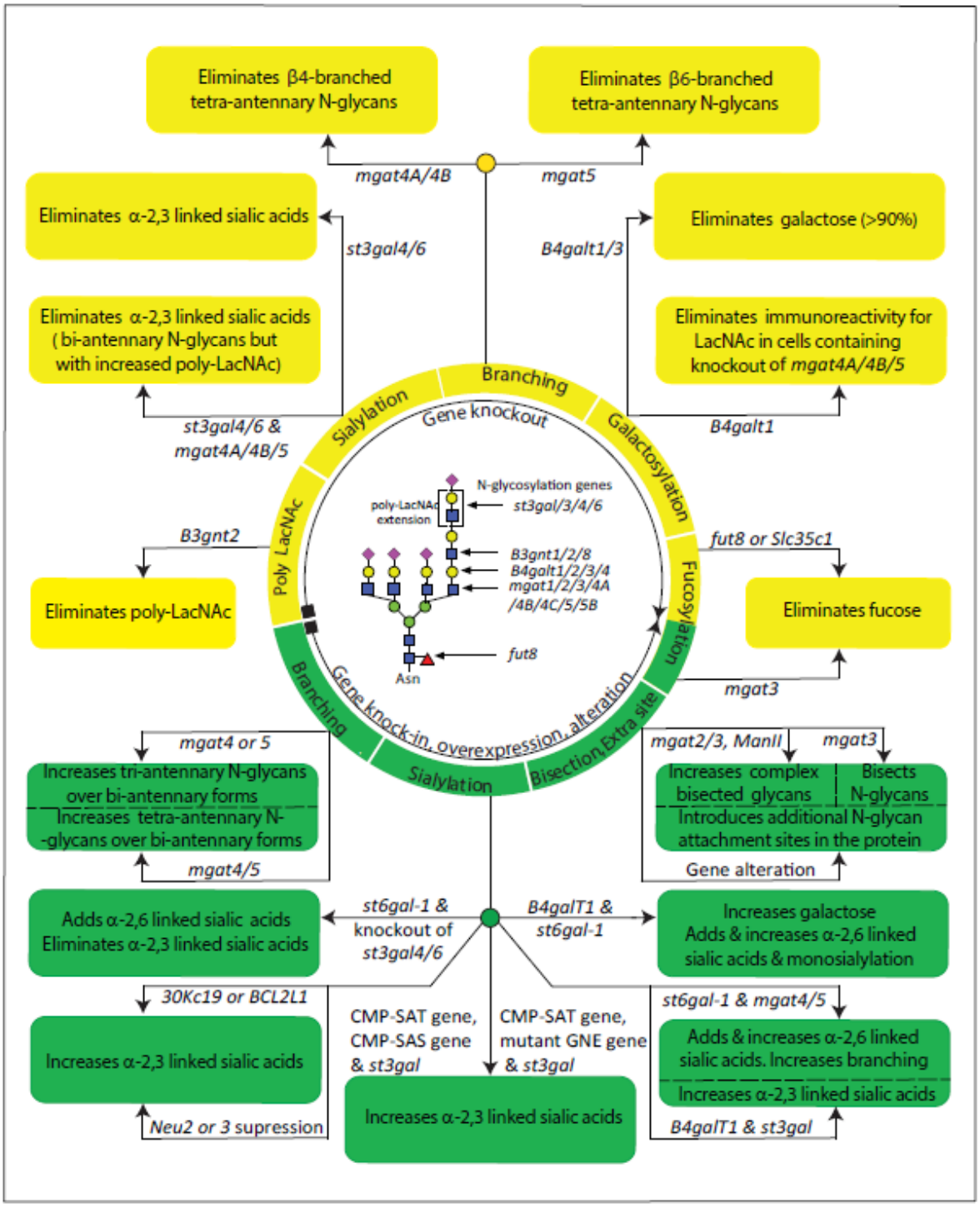

Figure 3

References

[1] J. Y. Kim, Y. G. Kim, and G. M. Lee, *Appl Microbiol Biotechnol*. **2012**, *93*, 917.
[2] G. Walsh, *Nat Biotechnol*. **2014**, *32*, 992.
[3] B. Kelley, *MAbs*. **2009**, *1*, 443.
[4] G. Walsh, *Nat Biotechnol*. **2010**, *28*, 917.
[5] R. J. Kaufman, L. C. Wasley, A. J. Spiliotes, S. D. Gossels, S. A. Latt, G. R. Larsen, and R. M. Kay, *Mol Cell Biol*. **1985**, *5*, 1750.
[6] M. Butler and A. Meneses-Acosta, *Appl Microbiol Biotechnol*. **2012**, *96*, 885.
[7] C. Morrison and R. Lähteenmäki, *Nat Biotechnol*. **2017**, *35*.
[8] C. Morrison, *Nat Biotechnol*. **2017**, *35*.
[9] G. Walsh, *Drug Discov Today*. **2010**, *15*, 773.
[10] B. A. Macher and U. Galili, *Biochim Biophys Acta*. **2008**, *1780*, 75.
[11] A. Varki, *Am J Phys Anthropol*. **2001**, *Suppl 33*, 54.
[12] A. Noguchi, C. J. Mukuria, E. Suzuki, and M. Naiki, *J Biochem*. **1995**, *117*, 59.
[13] C. J. Bosques, B. E. Collins, J. W. Meador, 3rd, H. Sarvaiya, J. L. Murphy, G. Dellorusso, D. A. Bulik, I. H. Hsu, N. Washburn, S. F. Sipsey, J. R. Myette, R. Raman, Z. Shriver, R. Sasisekharan, and G. Venkataraman, *Nat Biotechnol*. **2010**, *28*, 1153.
[14] E. U. Lee, J. Roth, and J. C. Paulson, *J Biol Chem*. **1989**, *264*, 13848.
[15] Y. Li and X. Chen, *Appl Microbiol Biotechnol*. **2012**, *94*, 887.
[16] C. Ferrara, P. Brunker, T. Suter, S. Moser, U. Puntener, and P. Umana, *Biotechnol Bioeng*. **2006**, *93*, 851.
[17] M. Kurogochi, M. Mori, K. Osumi, M. Tojino, S. Sugawara, S. Takashima, Y. Hirose, W. Tsukimura, M. Mizuno, J. Amano, A. Matsuda, M. Tomita, A. Takayanagi, S. Shoda, and T. Shirai, *PLoS One*. **2015**, *10*, e0132848.
[18] C. W. Lin, M. H. Tsai, S. T. Li, T. I. Tsai, K. C. Chu, Y. C. Liu, M. Y. Lai, C. Y. Wu, Y. C. Tseng, S. S. Shivatare, C. H. Wang, P. Chao, S. Y. Wang, H. W. Shih, Y. F. Zeng, T. H. You, J. Y. Liao, Y. C. Tu, Y. S. Lin, H. Y. Chuang, C. L. Chen, C. S. Tsai, C. C. Huang, N. H. Lin, C. Ma, C. Y. Wu, and C. H. Wong, *Proc Natl Acad Sci U S A*. **2015**, *112*, 10611.
[19] R. Jefferis, *Adv Exp Med Biol*. **2005**, *564*, 143.
[20] P. Umana, J. Jean-Mairet, R. Moudry, H. Amstutz, and J. E. Bailey, *Nat Biotechnol*. **1999**, *17*, 176.
[21] G. Ashwell and J. Harford, *Annu Rev Biochem*. **1982**, *51*, 531.
[22] D. Ghaderi, M. Zhang, N. Hurtado-Ziola, and A. Varki, *Biotechnol Genet Eng Rev*. **2012**, *28*, 147.
[23] D. Ghaderi, R. E. Taylor, V. Padler-Karavani, S. Diaz, and A. Varki, *Nat Biotechnol*. **2010**, *28*, 863.
[24] R. Jefferis, *J Immunol Res*. **2016**, *2016*, 5358272.
[25] A. Kuriakose, N. Chirmule, and P. Nair, *J Immunol Res*. **2016**, *2016*, 1298473.
[26] R. Jefferis, *Curr Pharm Biotechnol*. **2016**, *17*, 1333.
[27] S. Dietmair, L. K. Nielsen, and N. E. Timmins, *Biotechnol J*. **2012**, *7*, 75.
[28] P. Datta, R. J. Linhardt, and S. T. Sharfstein, *Biotechnol Bioeng*. **2013**, *110*, 1255.
[29] X. Xu, H. Nagarajan, N. E. Lewis, S. Pan, Z. Cai, X. Liu, W. Chen, M. Xie, W. Wang, S. Hammond, M. R. Andersen, N. Neff, B. Passarelli, W. Koh, H. C. Fan, J. Wang, Y. Gui, K. H. Lee, M. J. Betenbaugh, S. R. Quake, I. Famili, B. O. Palsson, and J. Wang, *Nat Biotechnol*. **2011**, *29*, 735.

[30] K. Brinkrolf, O. Rupp, H. Laux, F. Kollin, W. Ernst, B. Linke, R. Kofler, S. Romand, F. Hesse, W. E. Budach, S. Galosy, D. Muller, T. Noll, J. Wienberg, T. Jostock, M. Leonard, J. Grillari, A. Tauch, A. Goesmann, B. Helk, J. E. Mott, A. Puhler, and N. Borth, *Nat Biotechnol*. **2013**, *31*, 694.

[31] N. E. Lewis, X. Liu, Y. Li, H. Nagarajan, G. Yerganian, E. O'Brien, A. Bordbar, A. M. Roth, J. Rosenbloom, C. Bian, M. Xie, W. Chen, N. Li, D. Baycin-Hizal, H. Latif, J. Forster, M. J. Betenbaugh, I. Famili, X. Xu, J. Wang, and B. O. Palsson, *Nat Biotechnol*. **2013**, *31*, 759.

[32] E. J. Blondeel, R. Ho, S. Schulze, S. Sokolenko, S. R. Guillemette, I. Slivac, Y. Durocher, J. G. Guillemette, B. J. McConkey, D. Chang, and M. G. Aucoin, *J Biotechnol*. **2016**, *234*, 127.

[33] D. Ley, A. K. Seresht, M. Engmark, O. Magdenoska, K. F. Nielsen, H. F. Kildegaard, and M. R. Andersen, *Biotechnol Bioeng*. **2015**, *112*, 2373.

[34] A. Farrell, N. McLoughlin, J. J. Milne, I. W. Marison, and J. Bones, *J Proteome Res*. **2014**, *13*, 3144.

[35] H. F. Kildegaard, D. Baycin-Hizal, N. E. Lewis, and M. J. Betenbaugh, *Curr Opin Biotechnol*. **2013**, *24*, 1102.

[36] A. M. Lewis, N. R. Abu-Absi, M. C. Borys, and Z. J. Li, *Biotechnol Bioeng*. **2016**, *113*, 26.

[37] P. M. Rudd, R. J. Woods, M. R. Wormald, G. Opdenakker, A. K. Downing, I. D. Campbell, and R. A. Dwek, *Biochim Biophys Acta*. **1995**, *1248*, 1.

[38] R. B. Parekh, A. G. Tse, R. A. Dwek, A. F. Williams, and T. W. Rademacher, *Embo J* **1987**, *6*, 1233.

[39] A. F. Williams, R. B. Parekh, D. R. Wing, A. C. Willis, A. N. Barclay, R. Dalchau, J. W. Fabre, R. A. Dwek, and T. W. Rademacher, *Glycobiology*. **1993**, *3*, 339.

[40] G. Reuter and H. J. Gabius, *Cell Mol Life Sci*. **1999**, *55*, 398.

[41] F. J. Olson, M. Bäckström, H. Karlsson, J. Burchell, and G. C. Hansson, *Glycobiology*. **2005**, *15*, 177.

[42] C. Ravanat, C. Gachet, J. M. Herbert, S. Schuhler, J. C. Guillemot, F. Uzabiaga, C. Picard, P. Ferrara, M. Freund, and J. P. Cazenave, *Eur J Biochem*. **1994**, *223*, 203.

[43] J. Esko, K. Kimata, and U. Lindahl, in *Essentials of Glycobiology*, (Eds: A. Varki, R.D. Cummings, J.D. Esko, H.H. Freeze, P. Stanley, C.R. Bertozzi, G.W. Hart, and M.E. Etzler), Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY) **2009**, Ch. 16.

[44] Z. Yang, S. Wang, A. Halim, M. A. Schulz, M. Frodin, S. H. Rahman, M. B. Vester-Christensen, C. Behrens, C. Kristensen, S. Y. Vakhrushev, E. P. Bennett, H. H. Wandall, and H. Clausen, *Nat Biotechnol*. **2015**, *33*, 842.

[45] R. L. Shields, J. Lai, R. Keck, L. Y. O'Connell, K. Hong, Y. G. Meng, S. H. Weikert, and L. G. Presta, *J Biol Chem*. **2002**, *277*, 26733.

[46] S. Matsumiya, Y. Yamaguchi, J. Saito, M. Nagano, H. Sasakawa, S. Otaki, M. Satoh, K. Shitara, and K. Kato, *J Mol Biol*. **2007**, *368*, 767.

[47] A. M. Goetze, Y. D. Liu, Z. Zhang, B. Shah, E. Lee, P. V. Bondarenko, and G. C. Flynn, *Glycobiology*. **2011**, *21*, 949.

[48] T. Angata and A. Varki, *Chem Rev*. **2002**, *102*, 439.

[49] L. Santell, T. Ryll, T. Etcheverry, M. Santoris, G. Dutina, A. Wang, J. Gunson, and T. G. Warner, *Biochem Biophys Res Commun*. **1999**, *258*, 132.

[50] M. Takeuchi, S. Takasaki, H. Miyazaki, T. Kato, S. Hoshi, N. Kochibe, and A. Kobata, *J Biol Chem*. **1988**, *263*, 3657.

[51] M. Zhang, K. Koskie, J. S. Ross, K. J. Kayser, and M. V. Caple, *Biotechnol Bioeng*. **2010**, *105*, 1094.

[52] O. M. Pearce, H. Laubli, A. Verhagen, P. Secrest, J. Zhang, N. M. Varki, P. R. Crocker, J. D. Bui, and A. Varki, *Proc Natl Acad Sci U S A*. **2014**, *111*, 5998.

[53] V. Padler-Karavani, H. Yu, H. Cao, H. Chokhawala, F. Karp, N. Varki, X. Chen, and A. Varki, *Glycobiology*. **2008**, *18*, 818.

[54] M. Hedlund, V. Padler-Karavani, N. M. Varki, and A. Varki, *Proc Natl Acad Sci U S A*. **2008**, *105*, 18936.

[55] C. Yu, K. Gao, L. Zhu, W. Wang, L. Wang, F. Zhang, C. Liu, M. Li, M. R. Wormald, P. M. Rudd, and J. Wang, *Sci Rep*. **2016**, *7*, 20029.

[56] R. M. Anthony, F. Nimmerjahn, D. J. Ashline, V. N. Reinhold, J. C. Paulson, and J. V. Ravetch, *Science*. **2008**, *320*, 373.

[57] Y. Kaneko, F. Nimmerjahn, and J. V. Ravetch, *Science*. **2006**, *313*, 670.

[58] C. Y. Chung, Q. Wang, S. Yang, B. Yin, H. Zhang, and M. Betenbaugh, *Biotechnol J*. **2017**, *12*, 1600502.

[59] X. Yu, K. Baruah, D. J. Harvey, S. Vasiljevic, D. S. Alonzi, B. D. Song, M. K. Higgins, T. A. Bowden, C. N. Scanlan, and M. Crispin, *J Am Chem Soc*. **2013**, *135*, 9723.

[60] Y. Mimura, R. M. Kelly, L. Unwin, S. Albrecht, R. Jefferis, M. Goodall, Y. Mizukami, Y. Mimura-Kimura, T. Matsumoto, H. Ueoka, and P. M. Rudd, *J Immunol Methods*. **2016**, *428*, 30.

[61] C. Raymond, A. Robotham, M. Spearman, M. Butler, J. Kelly, and Y. Durocher, *MAbs*. **2015**, *7*, 571.

[62] Y. T. Jeong, O. Choi, H. R. Lim, Y. D. Son, H. J. Kim, and J. H. Kim, *J Microbiol Biotechnol*. **2008**, *18*, 1945.

[63] J. H. Lee, Y. G. Kim, and G. M. Lee, *Biotechnol Prog*. **2015**, *31*, 1133.

[64] Z. Wang, J. H. Park, H. H. Park, W. Tan, and T. H. Park, *Biotechnol Bioeng*. **2011**, *108*, 1634.

[65] A. Varki and R. Schauer, in *Essentials of Glycobiology*, (Eds: A. Varki, R.D. Cummings, J.D. Esko, H.H. Freeze, P. Stanley, C.R. Bertozzi, G.W. Hart, and M.E. Etzler), Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY) **2009**, Ch. 14.

[66] N. S. Wong, M. G. Yap, and D. I. Wang, *Biotechnol Bioeng*. **2006**, *93*, 1005.

[67] X. Gu and D. I. Wang, *Biotechnol Bioeng*. **1998**, *58*, 642.

[68] B. Yin, Q. Wang, C. Y. Chung, R. Bhattacharya, X. Ren, J. Tang, K. J. Yarema, and M. J. Betenbaugh, *Biotechnol Bioeng*. **2017**, *114*, 1899.

[69] Y. T. Jeong, O. Choi, Y. D. Son, S. Y. Park, and J. H. Kim, *Biotechnol Appl Biochem*. **2009**, *52*, 283.

[70] Y. D. Son, Y. T. Jeong, S. Y. Park, and J. H. Kim, *Glycobiology*. **2011**, *21*, 1019.

[71] E. Monti, A. Preti, B. Venerando, and G. Borsani, *Neurochem Res*. **2002**, *27*, 649.

[72] V. Seyrantepe, K. Landry, S. Trudel, J. A. Hassan, C. R. Morales, and A. V. Pshezhetsky, *J Biol Chem*. **2004**, *279*, 37021.

[73] Y. Kakugawa, T. Wada, K. Yamaguchi, H. Yamanami, K. Ouchi, I. Sato, and T. Miyagi, *Proc Natl Acad Sci U S A*. **2002**, *99*, 10718.

[74] A. Hinek, A. V. Pshezhetsky, M. von Itzstein, and B. Starcher, *J Biol Chem*. **2006**, *281*, 3698.

[75] N. de Geest, E. Bonten, L. Mann, J. de Sousa-Hitzler, C. Hahn, and A. d'Azzo, *Hum Mol Genet*. **2002**, *11*, 1455.

[76] M. J. Gramer, *Adv Biochem Eng Biotechnol*. **2014**, *139*, 123.

[77] J. Ferrari, J. Gunson, J. Lofgren, L. Krummen, and T. G. Warner, *Biotechnol Bioeng*. **1998**, *60*, 589.
[78] T. Shinkawa, K. Nakamura, N. Yamane, E. Shoji-Hosaka, Y. Kanda, M. Sakurada, K. Uchida, H. Anazawa, M. Satoh, M. Yamasaki, N. Hanai, and K. Shitara, *J Biol Chem*. **2003**, *278*, 3466.
[79] S. Krapp, Y. Mimura, R. Jefferis, R. Huber, and P. Sondermann, *J Mol Biol*. **2003**, *325*, 979.
[80] E. Suzuki, R. Niwa, S. Saji, M. Muta, M. Hirose, S. Iida, Y. Shiotsu, M. Satoh, K. Shitara, M. Kondo, and M. Toi, *Clin Cancer Res*. **2007**, *13*, 1875.
[81] T. T. Junttila, K. Parsons, C. Olsson, Y. Lu, Y. Xin, J. Theriault, L. Crocker, O. Pabonan, T. Baginski, G. Meng, K. Totpal, R. F. Kelley, and M. X. Sliwkowski, *Cancer Res*. **2010**, *70*, 4481.
[82] D. J. Becker and J. B. Lowe, *Glycobiology*. **2003**, *13*, 41r.
[83] Q. Wang, M. Stuczynski, Y. Gao, and M. J. Betenbaugh, *Methods Mol Biol*. **2015**, *1321*, 287.
[84] L. Malphettes, Y. Freyvert, J. Chang, P. Q. Liu, E. Chan, J. C. Miller, Z. Zhou, T. Nguyen, C. Tsai, A. W. Snowden, T. N. Collingwood, P. D. Gregory, and G. J. Cost, *Biotechnol Bioeng*. **2010**, *106*, 774.
[85] N. Yamane-Ohnuki, S. Kinoshita, M. Inoue-Urakubo, M. Kusunoki, S. Iida, R. Nakano, M. Wakitani, R. Niwa, M. Sakurada, K. Uchida, K. Shitara, and M. Satoh, *Biotechnol Bioeng*. **2004**, *87*, 614.
[86] K. F. Chan, W. Shahreel, C. Wan, G. Teo, N. Hayati, S. J. Tay, W. H. Tong, Y. Yang, P. M. Rudd, P. Zhang, and Z. Song, *Biotechnol J*. **2016**, *11*, 399.
[87] H. Schachter, *Biochem Cell Biol*. **1986**, *64*, 163.
[88] M. Schuster, P. Umana, C. Ferrara, P. Brunker, C. Gerdes, G. Waxenecker, S. Wiederkum, C. Schwager, H. Loibner, G. Himmler, and G. C. Mudde, *Cancer Res*. **2005**, *65*, 7934.
[89] A. M. Abdel Rahman, M. Ryczko, M. Nakano, J. Pawling, T. Rodrigues, A. Johswich, N. Taniguchi, and J. W. Dennis, *Glycobiology*. **2015**, *25*, 225.
[90] Y. Zhao, Y. Sato, T. Isaji, T. Fukuda, A. Matsumoto, E. Miyoshi, J. Gu, and N. Taniguchi, *FEBS J* **2008**, *275*, 1939.
[91] K. S. Lau, E. A. Partridge, A. Grigorian, C. I. Silvescu, V. N. Reinhold, M. Demetriou, and J. W. Dennis, *Cell*. **2007**, *129*, 123.
[92] J. W. Dennis, S. Laferte, C. Waghorne, M. L. Breitman, and R. S. Kerbel, *Science*. **1987**, *236*, 582.
[93] K. Shirato, K. Nakajima, H. Korekane, S. Takamatsu, C. Gao, T. Angata, K. Ohtsubo, and N. Taniguchi, *J Clin Biochem Nutr*. **2011**, *48*, 20.
[94] K. Fukuta, R. Abe, T. Yokomatsu, N. Kono, M. Asanagi, F. Omae, M. T. Minowa, M. Takeuchi, and T. Makino, *Glycobiology*. **2000**, *10*, 421.
[95] B. Yin, Y. Gao, C. Y. Chung, S. Yang, E. Blake, M. C. Stuczynski, J. Tang, H. F. Kildegaard, M. R. Andersen, H. Zhang, and M. J. Betenbaugh, *Biotechnol Bioeng*. **2015**, *112*, 2343.
[96] R. D. Marshall, *Biochem Soc Symp*. **1974**, 17.
[97] Y. Gavel and G. von Heijne, *Protein Eng*. **1990**, *3*, 433.
[98] H. A. Kaplan, J. K. Welply, and W. J. Lennarz, *Biochim Biophys Acta*. **1987**, *906*, 161.
[99] N. Ceaglio, M. Etcheverrigaray, R. Kratje, and M. Oggero, *Biochimie*. **2008**, *90*, 437.
[100] J. Jones, S. S. Krag, and M. J. Betenbaugh, *Biochim Biophys Acta*. **2005**, *1726*, 121.

[101] L. Kasturi, H. Chen, and S. H. Shakin-Eshleman, *Biochem J*. **1997**, *323 (Pt 2)*, 415.
[102] S. H. Shakin-Eshleman, S. L. Spitalnik, and L. Kasturi, *J Biol Chem*. **1996**, *271*, 6363.
[103] L. Kasturi, J. R. Eshleman, W. H. Wunner, and S. H. Shakin-Eshleman, *J Biol Chem*. **1995**, *270*, 14756.
[104] V. Picard, E. Ersdal-Badju, and S. C. Bock, *Biochemistry*. **1995**, *34*, 8433.
[105] E. Bause, H. Hettkamp, and G. Legler, *Biochem J*. **1982**, *203*, 761.
[106] J. L. Mellquist, L. Kasturi, S. L. Spitalnik, and S. H. Shakin-Eshleman, *Biochemistry*. **1998**, *37*, 6833.
[107] E. Bause, *Biochem Soc Trans*. **1984**, *12*, 514.
[108] E. Bause, *Biochem J*. **1983**, *209*, 331.
[109] S. Perlman, B. van den Hazel, J. Christiansen, S. Gram-Nielsen, C. B. Jeppesen, K. V. Andersen, T. Halkier, S. Okkels, and H. T. Schambye, *J Clin Endocrinol Metab*. **2003**, *88*, 3227.
[110] J. C. Egrie and J. K. Browne, *Br J Cancer*. **2001**, *84 Suppl 1*, 3.
[111] J. C. Egrie, E. Dwyer, J. K. Browne, A. Hitz, and M. A. Lykos, *Exp Hematol*. **2003**, *31*, 290.
[112] T. Chitlaru, C. Kronman, B. Velan, and A. Shafferman, *Biochem J*. **2001**, *354*, 613.
[113] T. Chitlaru, C. Kronman, B. Velan, and A. Shafferman, *Biochem J*. **2002**, *363*, 619.
[114] C. Kronman, T. Chitlaru, E. Elhanany, B. Velan, and A. Shafferman, *J Biol Chem*. **2000**, *275*, 29488.
[115] C. Kronman, B. Velan, D. Marcus, A. Ordentlich, S. Reuveny, and A. Shafferman, *Biochem J*. **1995**, *311 (Pt 3)*, 959.
[116] J. Powell and C. Gurk-Turner, *Proceedings (Baylor University. Medical Center)*. **2002**, *15*, 332.
[117] A. Gervais, Y. A. Hammel, S. Pelloux, P. Lepage, G. Baer, N. Carte, O. Sorokine, J. M. Strub, R. Koerner, E. Leize, and A. Van Dorsselaer, *Glycobiology*. **2003**, *13*, 179.
[118] E. Böhm, B. K. Seyfried, M. Dockal, M. Graninger, M. Hasslacher, M. Neurath, C. Konetschny, P. Matthiessen, A. Mitterer, and F. Scheiflinger, *BMC Biotechnology*. **2015**, *15*, 87.
[119] Y. Seo, G. M. Park, M. J. Oh, and H. J. An, *Bioanalysis*. **2017**, *9*, 1361.
[120] S. Houel, M. Hilliard, Y. Q. Yu, N. McLoughlin, S. M. Martin, P. M. Rudd, J. P. Williams, and W. Chen, *Anal Chem*. **2014**, *86*, 576.
[121] L. Zhu, Q. Guo, H. Guo, T. Liu, Y. Zheng, P. Gu, X. Chen, H. Wang, S. Hou, and Y. Guo, *MAbs*. **2014**, *6*, 1474.
[122] A. Seyhan and B. Ata, *International Journal of Women's Health*. **2011**, *3*, 243.
[123] C. Spahr, J. J. Kim, S. Deng, P. Kodama, Z. Xia, J. Tang, R. Zhang, S. Siu, N. Nuanmanee, B. Estes, J. Stevens, M. Zhou, and H. S. Lu, *Protein Science : A Publication of the Protein Society*. **2013**, *22*, 1739.
[124] J. Zaia, *Molecular & Cellular Proteomics*. **2013**, *12*, 885.
[125] P. L. DeAngelis, *Appl Microbiol Biotechnol*. **2012**, *94*, 295.
[126] U. Bhaskar, E. Sterner, A. M. Hickey, A. Onishi, F. Zhang, J. S. Dordick, and R. J. Linhardt, *Applied Microbiology and Biotechnology*. **2012**, *93*, 1.
[127] R. J. Linhardt, *Journal of Medicinal Chemistry*. **2003**, *46*, 2551.
[128] H. Liu, Z. Zhang, and R. J. Linhardt, *Natural Product Reports*. **2009**, *26*, 313.
[129] L. Borsig, in *Glycosaminoglycans in Development, Health and Disease*, Vol. 93 (Ed: L. Zhang), Elsevier **2010**, Ch. 335.
[130] B. Casu, A. Naggi, and G. Torri, *Matrix Biology*. **2010**, *29*, 442.

[131] A. Maraveyas, M. J. Johnson, Y. P. Xiao, and S. Noble, *Cancer and Metastasis Reviews*. **2010**, *29*, 777.
[132] F. Robert, *Journal of Hematology & Oncology*. **2010**, *3*, 3.
[133] L. Y. Yu, H. G. Garg, B. Li, R. J. Linhardt, and C. A. Hales, *Current Cancer Drug Targets*. **2010**, *10*, 229.
[134] P. L. Li, J. Z. Sheng, Y. H. Liu, J. Li, J. Liu, and F. S. Wang, *Medicinal Research Reviews*. **2013**, *33*, 665.
[135] N. Potdar, T. A. Gelbaya, J. C. Konje, and L. G. Nardo, *Human Reproduction Update*. **2013**, *19*, 674.
[136] S. Seshadri, S. K. Sunkara, Y. Khalaf, T. El-Toukhy, and H. Hamoda, *Reproductive Biomedicine Online*. **2012**, *25*, 572.
[137] M. D. Shastri, G. M. Peterson, N. Stewart, S. S. Sohal, and R. P. Patel, *Expert Opinion on Investigational Drugs*. **2014**, *23*, 357.
[138] V. K. Yadav, M. Saraswat, N. Chhikara, S. Singh, and S. Yadav, *Current Protein & Peptide Science*. **2013**, *14*, 61.
[139] J. Y. Baik, L. Gasimli, B. Yang, P. Datta, F. Zhang, C. A. Glass, J. D. Esko, R. J. Linhardt, and S. T. Sharfstein, *Metabolic Engineering*. **2012**, *14*, 81.
[140] P. Datta, G. Y. Li, B. Yang, X. Zhao, J. Y. Baik, T. R. Gemmill, S. T. Sharfstein, and R. J. Linhardt, *Journal of Biological Chemistry*. **2013**, *288*, 37308.
[141] G. Li, B. Yang, L. Li, F. Zhang, C. Xue, and R. J. Linhardt, *Analytical Biochemistry*. **2014**, *455*, 3.
[142] P. Hossler, B. C. Mulukutla, and W. S. Hu, *PLoS One*. **2007**, *2*, e713.
[143] F. J. Krambeck, S. V. Bennun, M. R. Andersen, and M. J. Betenbaugh, *PLoS One*. **2017**, *12*, e0175376.
[144] P. Umana and J. E. Bailey, *Biotechnol Bioeng*. **1997**, *55*, 890.
[145] H. Hefzi, K. S. Ang, M. Hanscho, A. Bordbar, D. Ruckerbauer, M. Lakshmanan, C. A. Orellana, D. Baycin-Hizal, Y. Huang, D. Ley, V. S. Martinez, S. Kyriakopoulos, N. E. Jimenez, D. C. Zielinski, L. E. Quek, T. Wulff, J. Arnsdorf, S. Li, J. S. Lee, G. Paglia, N. Loira, P. N. Spahn, L. E. Pedersen, J. M. Gutierrez, Z. A. King, A. M. Lund, H. Nagarajan, A. Thomas, A. M. Abdel-Haleem, J. Zanghellini, H. F. Kildegaard, B. G. Voldborg, Z. P. Gerdtzen, M. J. Betenbaugh, B. O. Palsson, M. R. Andersen, L. K. Nielsen, N. Borth, D. Y. Lee, and N. E. Lewis, *Cell Syst*. **2016**, *3*, 434.
[146] P. N. Spahn, A. H. Hansen, S. Kol, B. G. Voldborg, and N. E. Lewis, *Biotechnol J*. **2017**, *12*, 1600489.
[147] P. N. Spahn, A. H. Hansen, H. G. Hansen, J. Arnsdorf, H. F. Kildegaard, and N. E. Lewis, *Metab Eng*. **2016**, *33*, 52.
[148] I. Jimenez del Val, J. M. Nagy, and C. Kontoravdi, *Biotechnol Prog*. **2011**, *27*, 1730.
[149] F. J. Krambeck, S. V. Bennun, S. Narang, S. Choi, K. J. Yarema, and M. J. Betenbaugh, *Glycobiology*. **2009**, *19*, 1163.
[150] F. J. Krambeck and M. J. Betenbaugh, *Biotechnol Bioeng*. **2005**, *92*, 711.
[151] A. G. McDonald, J. M. Hayes, T. Bezak, S. A. Gluchowska, E. F. Cosgrave, W. B. Struwe, C. J. Stroop, H. Kok, T. van de Laar, P. M. Rudd, K. F. Tipton, and G. P. Davey, *J Cell Sci*. **2014**, *127*, 5014.
[152] S. R. Hamilton, P. Bobrowicz, B. Bobrowicz, R. C. Davidson, H. Li, T. Mitchell, J. H. Nett, S. Rausch, T. A. Stadheim, H. Wischnewski, S. Wildt, and T. U. Gerngross, *Science*. **2003**, *301*, 1244.

[153] S. J. North, H. H. Huang, S. Sundaram, J. Jang-Lee, A. T. Etienne, A. Trollope, S. Chalabi, A. Dell, P. Stanley, and S. M. Haslam, *J Biol Chem*. **2010**, *285*, 5759.
[154] K. N. Baker, M. H. Rendall, A. E. Hills, M. Hoare, R. B. Freedman, and D. C. James, *Biotechnol Bioeng*. **2001**, *73*, 188.
[155] D. Chee Furng Wong, K. Tin Kam Wong, L. Tang Goh, C. Kiat Heng, and M. Gek Sim Yap, *Biotechnol Bioeng*. **2005**, *89*, 164.
[156] M. Yang and M. Butler, *Cytotechnology*. **2000**, *34*, 83.
[157] N. Kochanowski, F. Blanchard, R. Cacan, F. Chirat, E. Guedon, A. Marc, and J. L. Goergen, *Biotechnol Bioeng*. **2008**, *100*, 721.
[158] N. S. Wong, L. Wati, P. M. Nissom, H. T. Feng, M. M. Lee, and M. G. Yap, *Biotechnol Bioeng*. **2010**, *107*, 321.
[159] P. M. Jedrzejewski, I. J. del Val, A. Constantinou, A. Dell, S. M. Haslam, K. M. Polizzi, and C. Kontoravdi, *Int J Mol Sci*. **2014**, *15*, 4492.
[160] J. Becker, M. Hackl, O. Rupp, T. Jakobi, J. Schneider, R. Szczepanowski, T. Bekel, N. Borth, A. Goesmann, and J. Grillari, *J Biotechnol*. **2011**, *156*, 227.
[161] D. Baycin-Hizal, D. L. Tabb, R. Chaerkady, L. Chen, N. E. Lewis, H. Nagarajan, V. Sarkaria, A. Kumar, D. Wolozny, J. Colao, E. Jacobson, Y. Tian, R. N. O'Meally, S. S. Krag, R. N. Cole, B. O. Palsson, H. Zhang, and M. Betenbaugh, *J Proteome Res*. **2012**, *11*, 5265.
[162] D. C. Frost and L. Li, in *Adv. Protein Chem. Struct. Biol*, Vol. 95 (Ed: R. Donev), Academic Press **2014**,
[163] S. B. Levery, C. Steentoft, A. Halim, Y. Narimatsu, H. Clausen, and S. Y. Vakhrushev, *Biochimica et Biophysica Acta (BBA) - General Subjects*. **2015**, *1850*, 33.
[164] H. Lu, Y. Zhang, and P. Yang, *National Science Review*. **2016**, *3*, 345.
[165] K. K. Palaniappan and C. R. Bertozzi, *Chem Rev*. **2016**, *116*, 14277.
[166] M. Thaysen-Andersen and N. H. Packer, *… et Biophysica Acta (BBA)-Proteins and …*. **2014**, *1844*, 1437.
[167] L. Veillon, S. Zhou, and Y. Mechref, *Methods in Enzymology*. **2017**, *585*, 431.
[168] Y. Yang, V. Franc, and A. Heck, *Trends in Biotechnology*. **2017**, *35*, 598.
[169] J. N. Liu, C. S. Jin, R. M. Cherian, N. G. Karlsson, and J. Holgersson, *J Biotechnol*. **2015**, *199*, 77.
[170] Z. Yang, A. Halim, Y. Narimatsu, H. J. Joshi, C. Steentoft, K. Schjoldager, M. A. Schulz, N. R. Sealover, K. J. Kayser, E. P. Bennett, S. B. Levery, S. Y. Vakhrushev, and H. Clausen, *Molecular & Cellular Proteomics*. **2014**, *13*, 3224.
[171] P. G. Slade, M. Hajivandi, C. M. Bartel, and S. F. Gorfien, *J Proteome Res*. **2012**, *11*, 6175.